



Formations Informatique > Big Data > Data Analysis > Hadoop

# Introduction à la plateforme Hadoop et à son écosystème

Référence VTBIGHA-EV

Durée du parcours (01h10)

Durée pédagogique (01h45)

Modalité et moyens pédagogique Vidéos - Exercices

#### À l'issue de cette formation vous serez capable de :

- Connaître les principes du framework Hadoop
- Utiliser la technologie MapReduce pour paralléliser des calculs sur des volumes importants de données.

#### Niveau requis

Avoir la connaissance d'un langage de programmation objet comme Java.

#### Public concerné

Chefs de projets, développeurs, data scientists et toute personne souhaitant comprendre les techniques de développement avec MapReduce dans l'environnement Hadoop.

Cette formation:

• Bénéficie d'un suivi de son exécution par un rapport de connexion.

Les temps donnés sont à titre indicatif, le stagiaire évoluant à son rythme.

## **Programme**

#### Origines d'Hadoop

- Qu'est-ce que le NoSQL? 01:34
- Définition du Big Data 01:48
- Histoire d'Hadoop 02:27

#### Installation d'un environnement Hadoop de base

- Écosystème complexe 01:45
- Distributions Hadoop 02:51
- Introduction à CDH: Cloudera Hadoop 01:09
- Démarrage d'une QuickStart VM 02:22

#### Stockage de fichiers : HDFS

- Présentation de l'HDFS 03:17
- Manipulation de fichiers en ligne de commande 04:01

#### Paradigme MapReduce

- Principe général 03:31
- Fonction Map 03:42
- Fonction Reduce 04:20



### Montée en compétences Conseil en financement Retour à l'emploi / Reskilling

#### Développement d'un premier MapReduce

- Mapper 06:52
- Reducer 02:41
- Lanceur 03:50
- Lancement de votre premier MapReduce 06:32
- Combineur 05:23

#### Écosystème Hadoop

- Hbase 05:14
- Hive 02:26
- Pig 01:18
- Zookeeper 04:04
- Sqoop 01:46
- Oozie 03:03
- Flume 02:18
- Kafka 02:08
- Spark 01:19